数理解析研究所講究録
625巻 1987年 21-30

対象集合によるネットワークデータベースの設計

九州大学工学部 古川哲也（Tetsuya Furukawa）
九州大学工学部 上林彌彦（Yahiko Kambayashi）

1. まえがき

データベースにおける重要な問題として、意味制約の保持と質問処理の効率化がある。ネットワークモデルに基づくデータベースではデータの対応をリンクで表しているめ、構造に制約があり、処理効率も構造によって異なる。意味制約としての従属性制約を構造に反映させると、冗長性が削減される。質問処理は一般に冗長性の付加により効率化される。従って、データベースの構造であるスキーマ設計では両方を考慮した最適なものを求める必要がある。

これまでのスキーマ設計は、実体関連モデルのような対応関係を表すものを用いて設計者の経験に基づいて行われていたが、体系的に設計を行う方法として次のものが考えられる。

(A) 意味制約を反映した非冗長なスキーマを質問処理効率化のために変換する。

(B) 質問処理のみを考慮したスキーマを意味制約の保持のために変換する。

意味制約による設計法として、関数従属性集合[1]，多値従属性集合[2]を用いたものがあり、質問処理効率化のためのスキーマの変換[3]もできる。 これらは(A)の方法に用いることができる。

本稿では、ネットワークスキーマ設計の問題点を明確にした後、(B) の方法について質問処理のための設計法として対象集合を用いたものを提案し、意味制約（従属性制約）を保持するためにスキーマを変換する方法を示す。この変換は従属性制約を対象として付加することによる。また、これら２つの方法の特徴を比較して最適なネットワークデータベースの設計法について検討する。

2. 基本的事項

ネットワークモデルの構造は、同じデータ項目（本稿では属性と呼ぶ）からな

るレコードの集合であるレコード型と、２つのレコード型間のレコードの１対多の対応を表す親子集合型の集合によって表現される。属性集合Xからなるレコード型RをR(X)で、親レコード型がR_o，子レコード型がR_mである親子集合型を$<R_o, R_m>$で表す。この構造はバックマン線図と呼ばれる有向グラフＢ(V,E) で表される。ここで、Vは各レコード型に対応する節点集合、Eは親レコード型から子レコード型へ向かう有向枝の集合である。値が定まればレコード型R(X)のレコードがただ１つ定まるような最小の属性集合KをRのキーといい、K(R)で表す。必ずしも$K(R) \subseteq X$である必要はなく、Rが子レコード型となる親子集合が定まれば（親レコード集合が定まれば）その子レコード間で$K(R) \cap X$ の値によってレコードが一意に定まればよい。バックマン線図では、キーを下線で、レコード型に含まれない属性は括弧を用いて表す。

バックマン線図Ｂの連結な部分グラフで表されるネットワーク構造をＢの部分ネットワーク構造という。特に属性集合Xと共通属性を持つレコード型とその間の経路のレコード型，親子集合型からなる部分ネットワーク構造をＢ(X)で表す。

本稿で対象とする質問は、指定された属性集合X_sの値に対応する属性集合X_Pの値を求めるというもので、Ｑ(X_s,X_P)で表す。X_sの値の指定、X_Pの値の出力は、それぞれ関係代数における選択，射影演算に対応している。ネットワークモデルにおける質問処理は、巡航操作によりレコードの対応関係を求めることに基づいており、これは自然結合演算に対応しているので、このクラスの質問は非常に一般的なものである。バックマン線図の部分グラフで表される質問Ｑの処理に必要なネットワーク構造、即ちＱの処理のために巡航が行われる部分をＱの質問スキーマB_Qと呼ぶ。B_QはＢ($X_s X_P$)の部分グラフである。

意味制約の保持では、従属性制約が数学的にも扱い易く、冗長性の除去からも重要である。本稿では特に関数従属性について議論する。ネットワークモデルにおける関数従属性は、次のように考えることができる。ネットワーク構造でのレコードの１つの対応をt、tの属性集合Xの値をt[X]とする。任意のレコードの２つの対応t_1,t_2について、$t_1[X] = t_2[X]$ならば$t_1[Y] = t_2[Y]$となるとき、関数従属性：$X \rightarrow Y$を満足している。従って、ネットワーク構造で常に満足される（ネットワーク構造が表す）関数従属性は、次のものである。

(a) １つのレコード型R(X)で、$K(R) \rightarrow X$。

(b) １つの親子集合型<$R_o(X_o)$,$R_m(X_m)$>で、$K(R_m) \to X_o$。
(c) (a),(b)の関数従属性集合から推移律により導かれるもの。

３．ネットワークスキーマ設計の問題点

ネットワークデータベースのスキーマ設計では、質問処理の効率化と意味制約の保持を考慮しなければならず、それらは次のようにまとめることができる。

データの一貫性を保つには、次の条件を満たす必要がある。[4)]

(1) 属性連結条件：任意の属性Aに対し、Ｂ(A)中の各レコード型R(X)について、$A \in X$。

これにより、任意の属性の対応は一致し、その属性の値が同じであるという関連を表す。

従属性制約についてネットワーク構造が満たすべき条件は、次のものとなる。与えられた関数従属性集合をS_{FD}とする。

(2) 関数従属性条件：バックマン線図で表現される関数従属性の集合は、S_{FD}と等価である。
(3) 適正キー条件：バックマン線図中の各レコード型R(X)に対し、K(R)は$K \to X$がS_{FD}から導ける最小の属性集合Kである。

質問$Q(X_s,X_p)$の処理効率の良さとして、次の基準を考える。[3)]
(a) X_sと共通属性を持たないレコード型を巡航せずにX_sの選択条件の検査ができる。
(b) X_s,X_Pと共通属性を持たないないレコード型を巡航しない。
(c) 得られた結果に重複がない。
これらから、質問スキーマB_Qが満たすべき条件を次のようにする。

(4) 選択属性連結条件：$B_Q(X_s)$中の各レコード型R(X)について、$X_s \cap X \neq \phi$。
(5) 質問属性連結条件：B_Q中の各レコード型R(X)について、$X_s X_P \cap X \neq \phi$。

(6) 解の非冗長性条件：B_Q中の各レコード型$R(X)$について、Bが表現する関数従属性により$X_S X_P \rightarrow K(R)$が導ける。

文献[3]の方法では、(4)〜(6)のうち指定された条件を満たすような質問スキーマが存在するようにネットワーク構造を変換できる。

ネットワークスキーマの設計では、これらの条件をなるべく満たし、冗長性の少ないものを求める必要がある。(1)や(2)を満たさないときは、データの更新の際にデータが一貫しているか、あるいは関数従属性を満足するかどうかの検査を必要とする。(3)は設計時にキーを計算できるが、親子集合型が1対多の対応であるため、キーとできない場合もでてくる。(4),(5),(6)が満足されないと質問処理のコストがかかる。これらの条件は、冗長性を犠牲にすることによりすべて満足されるが、冗長性に対しても管理が必要である。本稿の目的は、これらの条件をすべて満足し、冗長性ができるだけ少ないネットワーク構造を求めることにある。次の性質は、表現する従属性制約を変えないものであり、冗長性の除去に用いることができる。

(ア) 親子集合型$<R_o(X), R_m(Y)>$に対し、R_mを$X-Y \subseteq X' \subseteq XY$を満たす$R_m'(X')$で、$R_o$を$Y-(X-K(R_o)) \subseteq Y' \subseteq Y$を満たす$R_o'(Y')$で置き換えることができる。

(イ) レコード型$R_1(X_1), R_2(X_2)$に対し、$K(R_1)=K(R_2)$であればR_1とR_2のレコードの対応は１対１なので、２つのレコード型を１つにまとめて$R(X_1 X_2)$とすることができる。

(ア)は各レコード型を構成する属性集合の自由度によるもので、(イ)はレコード型の冗長性によるものである。

4. 質問処理のための構造

質問集合を効率良く処理できるネットワーク構造を設計する。意味的につながりのある属性集合を対象と呼ぶ。質問$Q(X_S, X_P)$では、属性集合$X_S X_P$の対応を求める必要があるので、$X_S X_P$を１つの対象とする。また、X_Sの選択条件は巡航操作で早く検査される方がよいので、X_Sも１つの対象とする。このようにして求めら

れた対象集合を用いてネットワーク構造を設計する。

［定義］ 質問集合S_Qに対する対象集合を$OBJ(S_Q)=\{X \mid X=X_S$又は$X=X_SX_P$, $Q(X_S,X_P)\in S_Q\}$とする。

［定義］ 対象集合OBJの共通集合演算に関する閉包OBJ^{i+}は、 任意の対象X_1,X_2 ($\in OBJ'$) に対し、 $X_1\cap X_2=X\ (\neq\phi)\in OBJ', OBJ\subseteq OBJ'$を満たす最小の対象集合OBJ'である。

Procedure 1：質問集合S_Qによるバックマン線図の作成
(1) $OBJ(S_Q)^{i+}$を求める。
(2) $OBJ(S_Q)^{i+}$の各要素Xを節点Xとする。
(3) 節点$X_1,X_2(X_1\subseteq X_2)$に対し、$X_1\subsetneq X_3\subsetneq X_2$となる節点$X_3$が存在しなければ有向枝$\langle R_1,R_2\rangle$を作る。
(4) 各節点Xをレコード型R(X)、各有向枝$\langle X_1,X_2\rangle$を親子集合型$\langle R_1(X_1),R_2(X_2)\rangle$を対応づける。一般に$R_i(X_i)$のキーは$X_i$となる。

(1)はSが共通集合演算で閉じるようにするためのもので、バックマン線図で同じ属性を含むレコード型をそれのみで連結とし、一貫性を保つために必要である。これにより条件(1)が満足される。B_Qは鎖構造となり、巡航操作はすべて有向枝の方向に（親レコード型から子レコード型に向かって）行われる。X_Sまでの節点で選択条件を満たすレコードの対応がただ１つ定まる。X_Sのそのレコードにつながる残りのレコードの対応をもとめればそれが解となる。

［定理１］ 質問集合S_Qに対し、Procedure 1により設計したネットワークスキーマでは、S_Q中の任意の質問に対して３節での条件(1),(4)～(6) を満足する鎖構造の質問スキーマが存在する。

証明） Procedure 1で作られたバックマン線図をBとする。任意の属性Aに対し、B(A)中のレコード型$R_1(X_1),R_2(X_2)(A\in X_i$, $i=1,2)$で、レコード型の属性集

合の集合は共通集合演算で閉じていることにより、Ｂ(A)中に$R(X)$($X = X_1 \cap X_2$)が存在する。RとR$_i$(i=1,2)を結ぶ経路のレコード型は、$X \subseteq X_i$，ステップ(3)より属性Xを含む。$A \in X$より属性Aを含み、条件(1)を満足する。

S_Q中の任意の質問$Q(X_S, X_P)$に対し、ステップ(2)よりＢ中にはX_S,$X_S X_P$からなるレコード型R_S,R_{SP}が存在し、ステップ(3)よりR_SはR_{SP}の祖先である。この鎖構造をＱの質問スキーマとすると条件(4),(5)を満足する。また、R_{SP}のキーは$X_S X_P$であり、条件(6)を満足する。

［例１］ 質問集合を$S_Q = \{Q_1(AD, BC), Q_2(DE, A)\}$とする。$OBJ(S_Q) = \{AD, A, ADE, ABCD\}$であり、$OBJ(S_Q) = OBJ(S_Q)^{i+}$である。 バックマン線図は図１となり$B_{Q1}$は$R_1, R_2, R_4$からなる部分ネットワーク構造、$B_{Q2}$は$R_1, R_2, R_3$からなる部分ネットワーク構造となる。

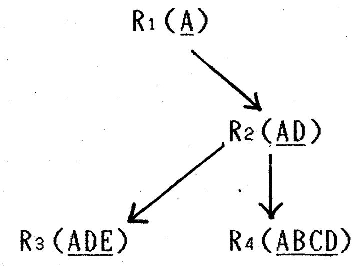


図１　質問処理のための構造

５．従属性制約の保持のための変換

４節で得られたネットワーク構造に関数従属性:$X \rightarrow Y$を反映させる。

［定義］ 属性集合Xの関数従属性集合S_{FD}に対する閉包Yは、$X \rightarrow Y$がS_{FD}から導かれるような最大の属性集合Yであり、X^{FD+}で表す。

［定義］ 関数従属性集合S_{FD}に対する対象集合を、$OBJ(S_{FD}) = \{X^{FD+} \mid X \rightarrow Y \in S_{FD}\}$とする。

Procedure 2：バックマン線図Ｂの関数従属性集合S_{FD}に対する変換

(1) $X \in OBJ(S_{FD})$に対し、 Ｂに属性集合Xからなるレコード型$R(X)$が存在しなければ$R(X)$を加える。

(2) (1)で加えた各レコード型$R(X)$に対し、 $X_1 \subseteq X$,$X \subseteq X_2$となる最大のX_1,最小のX_2のレコード型$R_1(X_1)$,$R_2(X_2)$について親子集合型$\langle R_1, R \rangle$,$\langle R, R_2 \rangle$を加える。

親子集合型<R_1,R_2>が存在すればそれを除く。

(3) (1)で加えたレコード型$R(X^{FD+})$のキーをXとする。

キー条件による冗長性の除去

(4) 各親子集合型<$R_o(X_o)$,$R_m(X_m)$>で、$K(R_o) \subsetneq X_m$であれば、$K(R_m)-\{X_o-K(R_o)\}$をR_mのキーとする。$K(R_m) \subsetneq X_o$であれば、$K(R_o)-\{X_m-K(R_m)\}$をR_oのキーとする。

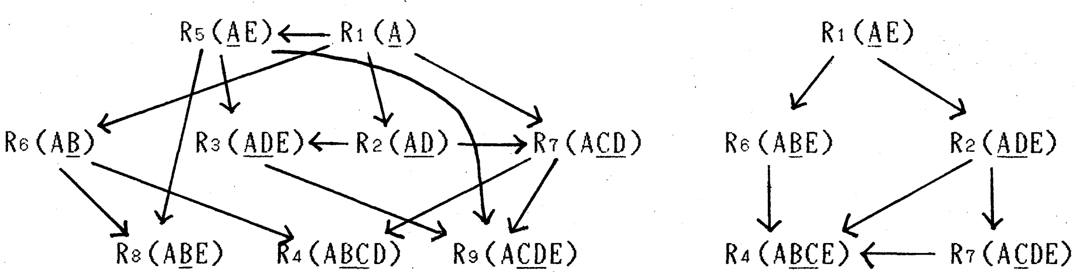
(5) レコード型$R_o(X_o)$,$R_m(X_m)$ ($R_o \neq R_m$) で、$K(R_o)=K(R_m)$であれば、R_o,R_mを除き$R(X_oX_m)$を加える。Rのキーを$K(R_o)$とする。親子集合型<R_o,R'>,<R_m,R'>又は<R',R_o>,<R',R_m>があれば、<R,R'>,<R',R>を加える。

(6) 各レコード型R_o,R_m($R_o \neq R_m$)で、$K(R_o) \subsetneq K(R_m)$であり、R_o,R_m間に枝<R_o,R_1>,<R_1,R_2>,…,<R_n,R_m>がなければ、枝<R_o,R_m>を加える。

［定理２］ Procedure 2はバックマン線図Ｂの条件(1),(4)～(6)に対する性質を変えずに、条件(2),(3)を満足するようにＢを変換する。

証明） レコード型の付加，併合、親子集合型の付加の方法から明らかに条件(1),(4)～(6)に対する性質を変えない。レコード型R(X)は、各関数従属性$X_1 \rightarrow X_2$に対しX_1^{FD+}との共通集合からなるレコード型を祖先に持ち、ステップ(4)よりその関数従属性をK(R)の計算に用いているので、K(R)は$K \rightarrow X$となる最小の属性集合Kである。また、明らかにキー条件，親子集合型で表現されている関数従属性は、S_{FD}に含まれるものであり、それ以外のものを表現していないし、S_{FD}中のすべての関数従属性を表現している。

［例２］ 関数従属性集合を$S_{FD}=\{A \rightarrow E, B \rightarrow A, C \rightarrow AD\}$を図１に反映する。 B→A



(a) (b)

図２ 関数従属性集合の反映

に対し、Bを定めるとA→EよりEも定まるのでB^{+}＝ABEである。 以下同様に、OBJ(S_{FD})＝{AE,ABE,ACDE}となる。ステップ(2),(3),(4)の結果は図２(a)となる。ここで、R_3のキーはR_4($\underline{A}$E)よりADとなり、R_4のキーはR_5,R_6よりBCである。 ステップ(5)では同じキーのレコード型が併合される。R_1とR_5,R_2とR_3,R_6とR_8,R_7とR_9が併合される。 ステップ(6)より親子集合型<R_7,R_4>が追加され、結果は図２(b)となる。

Procedure 2は、Procedure 1の結果のバックマン線図だけでなく、一般のものにも適用できる。ステップ(1),(2)はProcedure 1で同時に実行できる。その際、OBJ(S_Q)の各要素のキーをS_{FD}から求めることにより、レコード型の併合をこの段階で行うことができる。

［例３］ OBJ(S_Q)はキーを下線で表すと、{$\underline{A}$,$\underline{AD}$,$\underline{AD}$E,A$\underline{BC}$D}となる。OBJ(S_{FD})は{$\underline{A}$E,A$\underline{B}$E,A$\underline{C}$DE}であり、これらよりOBJ＝{$\underline{A}$E,$\underline{AD}$E,A$\underline{B}$E,A$\underline{BC}$D,A$\underline{C}$DE}が得られる。OBJ^{i+}（この例ではOBJに一致している）とキーの包含関係により、 図２(b)が得られる。

６．冗長性の除去

Procedure 2 の結果のバックマン線図では、冗長性が多い。３節での冗長性に対する(ｲ)の性質は、Procedure 2のステップ(5)で反映されているが、(ｱ)の冗長性が残る。質問集合に対する質問スキーマを考慮することにより、条件(1)～(6)の性質を変えることなく冗長性を削減することができる。

［定義］ 質問集合S_Qに対する質問スキーマを、$B_{SQ}(V,E)$ ($V=\cup V_i, E=\cup E_i$, $Q_i \in S_Q$とする。

Procedure 3：バックマン線図の冗長性の除去

レコード型R(X)に対し、<$R_o(X_o)$,R>が存在するX_oの和集合を$O_X(R)$,<R,$R_m(X_m)$>が存在するR_mの和集合を$M_X(R)$とする。

(1) B_{SQ}中の各レコード型R(X)で、B_{SQ}中に<$R_o(X_o)$,R>となるレコード型R_oが唯

一のとき、Rを$R'(X')$ $(X'=X-(X_o-M_x))$ で置き換える。

(2) B_{sq}外の各レコード型$R(X)$で、 $<R_o(X_o),R>$となるレコード型R_oに対し、Rを$R'(X')$ $(X'=X-(X_o-M_x))$ で置き換える。

(3) B_{sq}中の各レコード型$R(X)$で、 $<R,R_m(X_m)>$となるレコード型R_mに対し、Rを$R'(X')$ $(X'=X-\{X_m-\cup R_i-K(R)-O_x\})$ で置き換える。 R_iは$<R,R_i>$が存在するR_m以外のレコード型。

レコード型R(X)から除く属性が、冗長性に対する(ア)の性質よりも少ないのは、条件(1)～(6)に対する性質を変えないようにするためである。

［例４］ 図２のバックマン線図に対し、Procedure 3を適用すると図３となる。

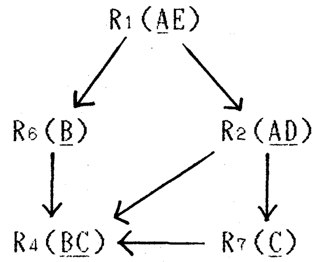


図３ 冗長性の除去

７．むすび

質問処理のための設計として対象集合からネットワークデータベースを設計する方法と、従属性制約保持のための変換を示した。本稿では、従属性制約として関数従属性のみを対象としたが、結合従属性にまで拡張可能である。この方法では、すべての質問は効率良く処理されるが、冗長性の問題，従属性制約の管理の問題がある。従属性制約を保持する構造に対する質問処理のための変換は、各質問に対し処理効率のクラスを選ぶことができ、冗長性との関係を考慮することができる。従って、ネットワークデータベースの設計は、次のように行うとよい。

１．特に効率良く処理できるようにしたい質問，構造に反映させたい従属性制約を対象集合とする構造を設計する。

２．代表的な質問に対し冗長性との兼ね合いを考えながらそれを効率よく処理できるように構造を変換する。

本稿での設計の結果では質問スキーマは鎖構造となるが、木構造となってもよ

く、効率が悪くなっても冗長性を減らし従属性制約を反映する変換があればさらによい設計ができる。

参考文献

1) Kuck,S.M. and Sagiv,Y., "Designing Globally Consistent Network Schemas," Proc. ACM SIGMOD, pp.185-195, May 1983.

2) Lien,Y.E., "On the Equivalence of Database Models," JACM, Vol.29, No.2, pp.333-362, April 1982.

3) Furukawa,T. and Kambayashi,Y., "Network Schema Conversion for Efficient Query Processing," Kyushu Univ. Thechnical Report, CSCE-86-C09, Dec. 1986.

4) Kambayashi,Y. and Furukawa,T., "Semantic Constrains Expressed by Netrwork Model," Proc. Int. Conf. on Foundations of Data Organization, pp.201-206, May 1985.

5) 古川，上林，"ネットワークモデルにおける木質問"，信学会技報，AL84-62, 1985年1月.

6) Ullman,J.D., Principles of Database Systems, 2nd Edition, Computer Science Press, 1983.

7) Maier,D., The Theorey of Relational Databases, Computer Science Press, 1983.